

Modeling Workload Impact in Multiple Unmanned Vehicle Supervisory Control

Birsen Donmez, Carl Nehme, and M. L. Cummings, *Member, IEEE*

Abstract—Discrete event simulations for futuristic unmanned vehicle (UV) systems enable a cost and time effective methodology for evaluating various autonomy and human-automation design parameters. Operator mental workload is an important factor to consider in such models. We present that the effects of operator workload on system performance can be modeled in such a simulation environment through a quantitative relation between operator attention and utilization, i.e., operator busy time used as a surrogate real-time workload measure. In order to validate our model, a heterogeneous UV simulation experiment was conducted with 74 participants. Performance-based measures of attention switching delays were incorporated in the discrete event simulation model via UV wait times due to operator attention inefficiencies (WTAI). Experimental results showed that WTAI is significantly associated with operator utilization (UT), such that high UT levels correspond to higher wait times. The inclusion of this empirical UT-WTAI relation in the discrete event simulation model of multiple UV supervisory control resulted in more accurate replications of data, as well as more accurate predictions for alternative UV team structures. These results have implications for the design of future human-UV systems, as well as more general multiple task supervisory control models.

Index Terms—Attention allocation, operator utilization, queuing theory, simulation, unmanned vehicles.

I. INTRODUCTION

SUPERVISORY control refers to intermittent operator interaction with a computer that closes an autonomous control loop [1]. With increased autonomy of unmanned vehicles (UVs), a human operator's role is shifting from controlling one vehicle to supervising multiple vehicles [2]. In the future, it is likely that a team of UVs will be composed of vehicles that vary in their capabilities or their assigned tasks, resulting in a "heterogeneous system" [3, 4]. Although the

appropriate size of a team is mission dependent, several experiments have all reached the same conclusion: There exists some upper bound to the number of vehicles that can be supervised by a single operator [5, 6]. To determine the most appropriate UV system architectures, it is critical to understand the impact of varying system design variables, such as level of vehicle autonomy, on the efficiency of human supervisory control of multiple UVs.

Human supervisory control is a complex system phenomenon with high levels of uncertainty, time-pressure, and a dynamically-changing environment. Discrete event simulation (DES), which models a system as a chronological sequence of events representing changes in system states [7], is particularly suited to model supervisory control systems due to their time-critical, event-driven nature. Such simulation models for futuristic systems allow for cost and time effective evaluation of different design parameters without conducting extensive experimentation, which is particularly critical in early conceptual design phases. While other modeling techniques, including agent-based models [8, 9], could potentially be used to capture human-UV interactions, a DES model was chosen as a first step, due to its ability to capture the temporal aspects of human-UV interactions. These temporal aspects determine important system limitations, and are defined below.

Using DES-based approaches, a few studies have attempted to computationally predict operator capacity when controlling multiple UVs [10-12], which generally focused on the use of neglect and interaction times to represent event and service rate distributions. Neglect time (NT) corresponds to the time that a robot or UV can be ignored before its performance drops below a predetermined threshold. Interaction time (IT) is defined as the amount of time the operator has to spend to bring a robot back to its peak performance. While these previous studies focused on clearly observable state transitions, the inherent delays that humans introduce in supervisory control systems were not considered. Vehicle wait times due to attention inefficiencies (WTAI) will occur as the operators fail to notice that the system needs their attention, and have been shown to significantly affect system performance [13].

This paper uses experimental data to demonstrate the need to incorporate human attention inefficiencies in models of human-UV systems, as well as a methodology to do so. As a first step, the effects of mental workload on UV operator attention inefficiencies are investigated. The relation between

Manuscript received November 26, 2008; revised July 22, 2009 and Jan 09, 2010. This work was funded by Charles River Analytics and the Office of Naval Research.

Birsen Donmez was a Postdoctoral Associate at the Humans and Automation Laboratory of Massachusetts Institute of Technology, Cambridge, MA 02139 USA. She is now an Assistant Professor at the University of Toronto Mechanical and Industrial Engineering Department, Toronto, ON M5S 3G8 Canada; phone: 416-978-7399; fax: 416-978-7753; e-mail: donmez@mie.utoronto.ca.

Carl Nehme is a consultant with McKinsey & Company, Dubai; e-mail: nehme@mit.edu.

M. L. Cummings is an Associate Professor at the Massachusetts Institute of Technology Aeronautics and Astronautics Department, Cambridge, MA 02139 USA; e-mail: missyc@mit.edu.

operator utilization (UT), a surrogate measure of workload, and a performance-based measure of inattention is incorporated into a DES model of multiple UV supervisory control through WTAI. Without the inclusion of the UT-WTAI relation, the DES model fails to provide accurate replication and prediction of the observed data.

II. BACKGROUND

Discrete event simulations are based on queuing theory, which model the human as a single server serially attending the arrival of events [14-16]. These models can also be extended to represent operator parallel processing through the introduction of multiple servers [17, 18]. In addition to the application of discrete event simulations to operator control of multiple robots as discussed previously, they have also been successfully applied to other supervisory control domains such as air traffic control [19]. However, as previously mentioned, the existing models of multiple robot control did not explicitly include operator cognitive inefficiencies, either as an input or an output.

A primary limiting factor in single operator-multiple UV systems is operator workload. Indeed, this limitation on the control of multiple UVs extends to any supervisory control task requiring divided attention across multiple tasks such as air traffic control and even supervisors multi-tasking in a command center like an air operation center.

Mental workload results from the demands a task imposes on the operator's limited resources; it is fundamentally determined by the relationship between resource supply and task demand [14]. While there are a number of different ways to measure workload [20, 21], given the temporal nature of supervisory control systems, particularly those in multi-UV control, we use utilization as a proxy for measuring mental workload. Utilization is a term found in systems engineering settings and refers to the "percent busy time" of an operator, i.e., given a time period, what percentage of time that person was busy. In supervisory control settings, this is generally meant as the time an operator is directed by external events to complete a task (e.g., replanning the path of a UV because of an emergent target). Compared to more common measures of workload (e.g., pupil dilation, NASA TLX), utilization provides the means to assess workload in real-time, non-intrusively. What is not included in this measurement is the time spent monitoring the system, i.e., just watching the displays and/or waiting for something to happen. While arguably this is not a perfect measure of mental workload, another strength of such a measure is its ratio scale, which allows it to be used in quantitative models.

The concept of utilization as a mental workload measure has been used in numerous studies examining supervisory controller performance [19, 22, 23]. These studies generally support that when tasked beyond 70% utilization, operators' performances decline. In terms of operator attention, high levels of arousal have been shown to induce perceptual narrowing [24].

While not well established empirically, there is some reason

to anticipate a decrease in performance with low levels as well as high levels of utilization. Previous research suggests an inverted-U shape between arousal/workload level and performance [25-27], indicating a decrease in performance with both low and high levels of arousal, which can occur as a function of utilization. As for attention, it has been established that vigilance decrement occurs when low arousal is experienced for extended periods of time [28].

Given the previous research showing that supervisory control performance drops when utilization is greater than 70%, and that there might be performance declines at high and low levels of utilization, we investigated whether the relationship between utilization and performance could be used to not just describe observed human behavior, but also be used to predict it. However, rather than connecting workload directly to performance, we captured effects of workload through delays introduced to the system by humans, which is more appropriate for incorporation to a DES model of human-UV systems.

Previous queuing theory based human information processing models have also used server utilization as a way to model workload [29, 30]. Although these models have successfully predicted workload and performance, the level of information processing detail captured was at the perceptual level and the human was represented by multiple servers. Supervisory control of complex systems requires operators to handle high-level tasking through reasoning and judgment, and these tasks are better fit to model at a higher (more abstract/cognitive) level as will be discussed later. Assessing the relation between workload and performance provides a parsimonious way to incorporate workload effects in high level models of human-system interaction when detailed level information processing models are not available.

In order to address the general explicit lack of accounting for human cognitive inefficiencies in models of human-UV interactions, this paper presents a queuing theory-based discrete event simulation model of a single operator supervising multiple heterogeneous UVs that includes a utilization-attention inefficiency component.

III. DISCRETE EVENT SIMULATION MODEL OF UV SUPERVISORY CONTROL

The proposed model utilizes queuing theory to build a discrete event simulation model by capitalizing on the event-driven nature of human-supervisory control (Fig. 1). This section presents an overview of the DES model and the details relevant to the focus of this paper. Further details can be found in [31].

The human operator, responsible for multiple UV supervision, is modeled as a server in a queuing system with discrete events representing both endogenous and exogenous situations an operator must address. Endogenous events, which are vehicle-generated or operator induced, are events created internally within the supervisory control system, such as when an operator elects to re-plan an existing UV path in order to reach a goal in a shorter time. It is important to note

that this interaction may not be required by the system and can therefore be operator-induced. Events which result from unexpected external environmental conditions that create the need for operator interaction are defined as exogenous events, such as emergent threat areas which require re-planning vehicle trajectories.

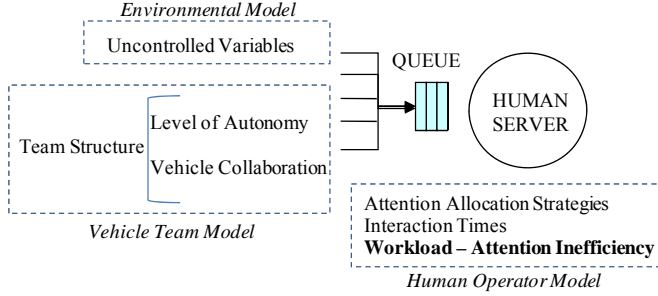


Fig 1. A high level representation of the discrete event simulation model

The design variables that serve as inputs to the model in Fig. 1 are composed of variables related to the vehicle team (team structure, level of autonomy, and vehicle collaboration), the human operator (interaction times, operator attention allocation strategies, and operator workload/attention inefficiency), and a model of environment unpredictability. These are discussed below in further detail.

A. Vehicle Team Input Variables

Team structure represents the number and type of vehicles included in the system. By representing each vehicle through a distinct input stream, the model is able to capture heterogeneous team composition since it includes different arrival processes for events associated with different vehicles. This is similar to the previously-discussed concept of neglect time except that in this case, the neglect time is for a specific event and not for the whole vehicle (i.e., other events associated with the same vehicle can still be generated while a specific event type is being neglected). Because NTs represent the time a vehicle can operate without human intervention, they effectively represent degrees of autonomy, i.e., the longer the NTs, the more autonomous the vehicle. Lastly, the model captures the effect of vehicle collaboration by taking into account the effect of servicing a particular event belonging to one vehicle on the arrival process of another event belonging to another vehicle. The types of vehicle autonomy and vehicle collaboration are influenced by the team structure, i.e., the different vehicle types.

B. Human Operator Input Variables

Our DES model represents the human server at a high level, capturing human performance holistically and stochastically through event service times (measured as the time from when operators engage a task to when they finish it), attention allocation strategies (i.e., strategies in choosing what task to service next), and attention inefficiencies.

The length of time it takes the operator to deal with an event, also known as interaction time, is captured through a

distribution of event service times. Interaction times occur for a single vehicle task, so in order to model the effect of an operator controlling multiple vehicles, the model should consider how and when operators elect to attend to the vehicles, also known as attention allocation [32].

The model in Fig. 1 captures two attention allocation strategies that can impact the effectiveness of human-UV interaction. The first strategy is the amount of operator-initiated re-planning. Since this model supports endogenous events that are both vehicle-generated and operator-induced, the rate at which operator-induced events arrive to the system depends on the operator's desire to interact with the vehicles beyond unavoidable vehicle-generated events.

Second, the queuing policy that is used by the operator, i.e., which task waiting in a queue the operator elects to service, is also represented. Examples include the first-in-first-out (FIFO) queuing scheme as well as the highest attribute first (HAF) strategy [33]. The HAF strategy is similar to a preemptive priority scheme in that high priority events are serviced first, except that there is no preemption. Therefore, if an event is generated with a priority higher than any of the events already in the system, it will be moved to the front of the queue but will not preempt a lower priority vehicle that is already being serviced.

The need for the last human operator input, the workload-attention inefficiency component, is the hypothesis of this paper and will be discussed in more detail in a subsequent section.

C. Arrival Events

As discussed previously, there are two categories of arrival events, endogenous and exogenous. Endogenous events are those created either by a vehicle (e.g., a UV requests permission from an operator to move to the next target), or by a human (a human initiating a new route without prompting by the system). Since an endogenous event associated with a specific vehicle generally requires attention before an event of the same type can be generated, the arrival process is one of correlated arrivals. For example, if an operator elects to initiate a path re-plan for vehicle A, he or she must finish servicing that event before electing to re-plan vehicle A's path again. In order to model this phenomenon, the model uses a closed queuing network paradigm such that each endogenous event type in the system (where each endogenous event type is associated with a specific vehicle) has a population of one [34]. The arrival process can therefore be described by a probabilistic distribution over a random variable, which represents the time between the completion of a service for an event associated with a specific vehicle and the next event being generated by that vehicle.

Exogenous events stem from sources external to the vehicle (weather, enemy movements, etc.). For example, many emergent threats can arise simultaneously, each requiring operator intervention, thus creating a queue. Therefore, the arrival process in the case of exogenous events is generally one of independent arrivals. The arrival process can therefore be described by a probabilistic distribution over a random

variable which represents the inter-arrival times between exogenous events.

In order to capture interaction between different event types, the servicing of one event type can be modeled to have an effect on the arrival process of another. For example, a UV might be modeled through two event types; a) the need for operator interaction whenever the operator is required to identify a target, and b) the need for operator interaction once target identification is complete and the vehicle requires a new assignment. In this case, event type (b) is generated only after event type (a) is serviced by the operator.

By modeling the operator as a single server, this model assumes serial operator interaction, such that events arriving while the operator is busy will wait in a queue. Although, it is possible for humans to multi-task, the appropriateness of the human model depends on the level of processing detail under consideration. When considering supervisory control tasks for complex systems such as those in UV systems, humans are generally required to handle high level tasking that involves application of human judgment and reasoning. While operators can rapidly switch between cognitive tasks, any sequence of tasks requiring complex cognition will form a queue and consequently, wait times will build [13]. As such, humans will act as serial processors in that they can only solve a single complex task at a time [35, 36]. While the serial processing model has been applied to capture higher level tasking [19], the parallel processing has been more generally applied for capturing lower-level perceptual processing such as that involved in driving tasks [18, 37]. For example, Schmidt [19] has suggested a single server queuing system as appropriate for modeling an air traffic controller in charge of conflict assessment and resolution, as well as general routing type tasks.

Based on the assumption of serial operator interaction, the service processes can be described by probabilistic distributions representing the interval from the time the operator decides to service a particular event until the operator is done servicing (this applies to both endogenous and exogenous events).

IV. EXPERIMENTAL MOTIVATION FOR A DES WORKLOAD MODEL

An online experiment was conducted to validate the previously described DES model for different vehicle team structures, with the ultimate goal to predict performance of various human-UV systems. Prior to this experiment, there was no workload-attention inefficiency component in the DES model. Moreover, the initial model validation reported in [31] was conducted on experimental data from sixteen participants. Given the relatively small sample size of the previous validation experiment, we wanted to ensure that any significant trends resulting from the current experiment had higher statistical power. Thus, online data collection was chosen as a means to increase our sample size. However, in order to decrease the likelihood of participant withdrawal from the online experiment, we kept the trials fairly short at 10

minutes. To ensure the validity of online experimentation, a pilot study was conducted with 15 participants completing the experiment online, and an additional 15 completing it in a laboratory setting with an experimenter present. No significant differences were observed between the two groups for the variables of interest [38].

As will be shown through the experimental results, the model without considering a workload-attention relationship does not adequately replicate results of the human-in-the-loop trials. Model inaccuracies provided the motivation to incorporate the human delays in the DES model.

A. Participants

Seventy-four participants, 6 females and 68 males between the ages of 18-50 completed the study. There were 36 participants between the ages of 18 and 25, 30 participants between 26 and 35, and eight participants whose age was greater than 35. The majority of participants were students and some were UV researchers from industry ($n=14$). Participants were randomly assigned to the experimental conditions based on the order they logged in to the online server. The breakdown of UV researchers across different experimental conditions was fairly constant ($n = 4, 5, 5$). There was no monetary compensation for participation; however the best performer received a \$200 gift certificate.

B. Experimental Test-bed

The Research Environment for Supervisory Control of Heterogeneous Unmanned Vehicles (RESCHU) simulator was used in the experiment, and allowed operators to control a team of UVs composed of unmanned air and underwater vehicles (UAVs and UUVs). All vehicles were engaged in surveillance tasks, with the ultimate mission of locating specific objects of interest in urban coastal and inland settings. While there was only a single UUV type, there were two UAV types, one that provided high level sensor coverage (akin to a Global Hawk UAV), while the other provided more low-level target surveillance and video gathering (similar to a Predator UAV). Thus, there were three different vehicle types under control for a single operator. Because previous research has shown that to allow for the simultaneous supervision and payload management (e.g., managing cameras for target identification) of multiple unmanned vehicles, navigation tasks for the different vehicles should be highly automated [6], this was a basic assumption for this simulation.

The RESCHU interface consisted of five major sections (Fig. 2). The map displayed the locations of vehicles, threat areas, and areas of interests (AOIs) (Fig. 3a). Vehicle control was carried out on the map, such as changing vehicle paths, adding a waypoint (a destination along the path), or assigning an AOI to a vehicle. The main events in the mission (i.e., vehicles arriving to goals, or automatic assignment to new targets) were displayed in the message box, along with a timestamp (Fig. 3b). When the vehicles reached an AOI, a simulated video feed was displayed in the camera window. The participant had to visually identify a target in this simulated video feed. Example targets and objects of interest

included cars, swimming pools, helipads, etc.

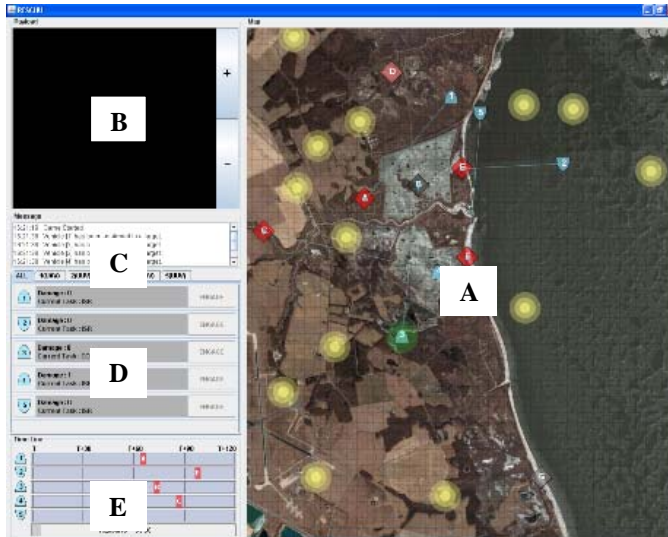


Fig 2. RESCHU interface (A: map, B: camera window, C: message box, D: control panel, E: timeline)

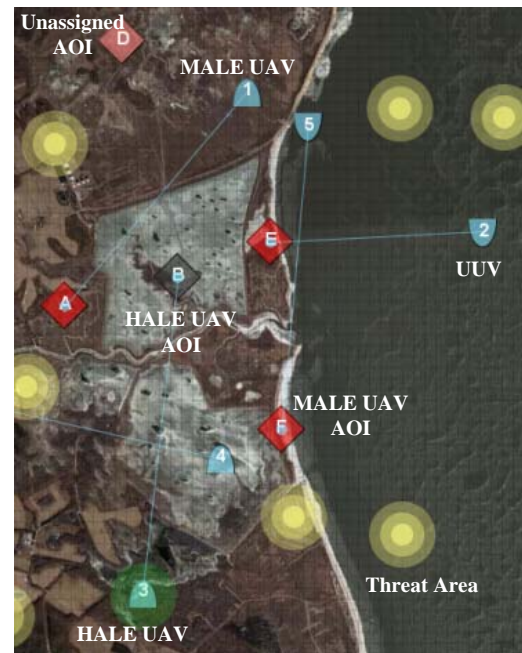
The control panel provided vehicle health information, as well as information on the vehicle's mission (Fig. 3c). The timeline displayed the estimated time of arrival to waypoints and AOIs. Beneath the timeline was a mission progress bar that showed the amount of time remaining in the total simulation.

As discussed previously, three types of vehicles were used in this experiment: a high altitude long endurance (HALE) UAV, medium altitude long endurance (MALE) UAVs, and UUVs. Both the MALE UAVs and the UUVs traveled to areas of interest (red AOIs in Fig. 3a) with a pre-determined target that needed to be visually acquired by the operator. MALE UAVs could travel to any AOI (both shore and land), whereas UUVs could only travel to AOIs that were on the shoreline. A HALE UAV traveled to AOIs that did not yet have a target specified (grey AOIs in Fig. 3a), and carried a Synthetic Aperture Radar (SAR)-type sensor, which allowed for target specification. These newly discovered targets were later acquired by a MALE UAV or a UUV.

When the vehicles completed their assigned tasking, an automated-path planner automatically assigned the HALE UAV to an AOI that needed intelligence, and the MALE UAVs and UUVs to AOIs with pre-determined targets. The automatically-assigned AOIs were not necessarily the optimal choice. The operator could change the assigned AOI, and could avoid threat areas by changing a vehicle's goal or adding a waypoint to the path of the vehicle in order to go around the threat area.

When a vehicle arrived to an AOI, a visual flashing alert indicated that the operator could engage the payload. For a HALE UAV, clicking the engage button resulted in the uncovering of the target in the AOI. For a MALE UAV or a UUV, engaging the payload caused the camera window to display the simulated live video feed (Fig. 3b). The operator then had to complete a search task by panning and zooming

the camera until the specified target was located. Once the operator submitted the target identification, the message box notified the operator on the accuracy of response (used to simulate feedback that real operators get from their commanders or teammates as a consequence of their actions), and the vehicle was automatically re-assigned to a new AOI.



(a)



(b)



(c)

Fig 3. A section of the map (b) Activated camera view and message box (c) Control panel and timeline

Participants were instructed to maximize their score by 1) avoiding threat areas that dynamically changed, 2) completing as many of the search tasks correctly, 3) taking advantage of re-planning when possible to minimize vehicle travel times between AOIs, 4) ensuring a vehicle was always assigned to an AOI whenever possible.

The UVs were not modeled on real UV performance data as this experiment simulated a futuristic system, i.e., there are no operational command and control systems with integrated heterogeneous unmanned operations. However, to create some realism, UUVs were modeled to move slower than UAVs, based on typical current platform capabilities. In the

experiment, the UVs required human intervention multiple times, creating a fast-paced scenario, and thus represented high workload situations.

C. Experimental Design and Independent Variables

The experiment was a completely randomized design with vehicle team heterogeneity level as a between-subject condition: none ($n=26$), medium ($n=25$), and high ($n=23$). The no heterogeneity condition included five MALE UAVs. The medium heterogeneous level had three MALE UAVs and two UUVs. Because the UUVs were slower than UAVs, they produced events less frequently. The maximum level of heterogeneity required managing two MALE UAVs, two UUVs, and one HALE UAV. HALE UAVs were restricted to grey AOIs, which appeared at a ratio of five-to-two, as compared to red AOIs, which the UUVs and MALE UAVs could visit without assistance from the HALE. Thus, the arrival rates of events for HALE UAVs were different than both the MALE UAVs and UUVs. Moreover, service times were different since the HALE UAVs required just milliseconds of service time (operators clicking the engage button). Lastly, because the UUVs were slower than UAVs and the HALE UAVs did not have an associated visual task, the no heterogeneity condition composed of five MALE UAVs was the highest tempo scenario, followed by the medium and then the high heterogeneity conditions.

D. Procedure

The online experiment began with an interactive tutorial followed by an open-ended practice session. The interactive tutorial had to be completed before the participants could start the practice session. During the interactive tutorial, the participants had to repeat a task until they performed it correctly. Thus, a major part of the training took place during the interactive tutorial. After participants felt comfortable with the task and the interface, they could end the practice session and start the 10 minute experimental session. Pilot participants were observed to spend on average 10 minutes doing the practice session. The website was password protected and participation was via invitation only. All data were recorded to an online database. Demographic information was collected via a questionnaire presented before the tutorial.

V. DES REPLICATIONS WITHOUT MODELING WORKLOAD EFFECTS

This section presents the results obtained from the experiment, followed by the analysis of the model's ability to describe the observed data and predict how changes in the vehicle heterogeneity structure will alter variables of interest.

The variables of interest for evaluating model predictions were score, average search task wait time, and operator utilization. Mission performance was assessed via score, which was calculated as the proportion of the number of targets correctly identified normalized by the number of all possible targets that could have been identified. Search task wait time was calculated from vehicle arrival to an AOI and

the operator engaging the search task. Average search task wait time assessed system performance efficiency since it demonstrated the effects of operator inefficiencies via system delays. Operator utilization was calculated as the proportion of time the operator actively interacted with the display (e.g., adding a way point, engaging in a visual task, etc.) during the course of the experiment. Utilization therefore excluded any monitoring time expended by operators.

A. Observed Effects

A preliminary analysis demonstrated significant correlations between the three variables of interest: utilization/score ($\rho=-.25$, $p=.03$), utilization/search task wait times ($\rho=-.50$, $p<.0001$), and score/search task wait times ($\rho=-.58$, $p<.0001$). Because these three measures are correlated, Multivariate Analysis of Variance (MANOVA) was performed to control for the inflation of Type I error. Significant findings were followed with univariate analysis to assess the magnitude of the effect that vehicle heterogeneity level had on each variable.

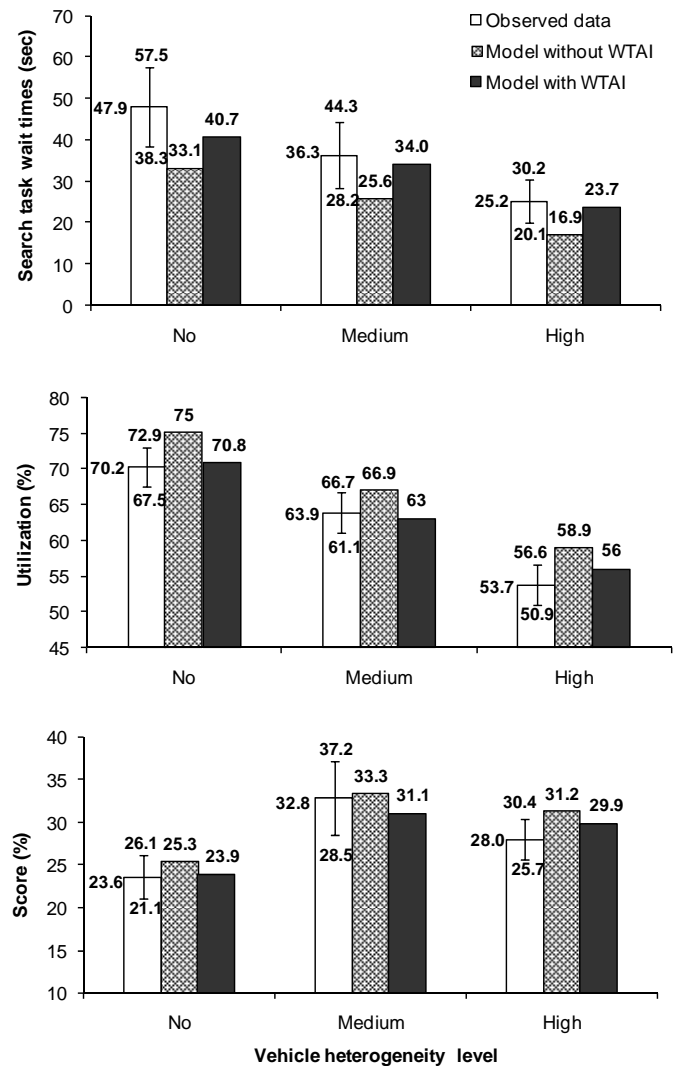


Fig 4. Observed data with 95% confidence intervals and model replications for the three dependent variables of interest: (a) search task wait times, (b) utilization, and (c) score.

The MANOVA results indicated that there were significant effects of heterogeneity level (Wilks' Lambda=0.4, $F(6,138)=13.33$, $p<.0001$). The univariate analysis suggested that the effect of heterogeneity level is attributable to the differences observed in all of the three variables of interest (Fig. 4). There were significant differences between different heterogeneity levels for utilization ($F(2,71)=33.31$, $p<.0001$), score ($F(2,71)=8.13$, $p=.0007$), and search task wait times ($F(2,71)=7.75$, $p=.0009$).

The pair-wise comparisons (Table 1) revealed that, as expected due to the tempo of arriving events, utilization was the highest for the no heterogeneity level, followed by the medium and the high heterogeneity levels. UUVs, which spent a considerable amount of time underwater, required less frequent interaction with the human operator than UAVs. Additionally, HALE UAVs required shorter interactions than the MALE UAVs. Thus, as the level of heterogeneity increased, operators interacted less frequently with the vehicles due to longer neglect times and shorter interaction periods. This cascading effect was also seen in the wait time metric as the homogenous (no heterogeneity) team structure generated significantly longer search task wait times as compared to both the medium and the high heterogeneity team structures. Because the operators had to interact more often with the MALES, the no heterogeneity service queues were larger, thus generating longer wait times.

TABLE I
PAIR-WISE COMPARISONS FOR DEPENDENT MEASURES

Comparison of heterogeneity levels	Difference	t-value (df: 71)	p	95% CI
<i>Search task wait times</i>				
No vs. medium	11.63 s	2.06	.04	(0.35, 22.92)
No vs. high	22.74 s	3.93	.0002	(11.21, 34.27)
Medium vs. high	11.11 s	1.90	.06	(-0.53, 22.75)
<i>Utilization</i>				
No vs. medium	6.29 %	3.17	.002	(2.34, 10.25)
No vs. high	16.45 %	8.12	<.0001	(12.41, 20.49)
Medium vs. high	10.16 %	4.97	<.0001	(6.08, 14.24)
<i>Score</i>				
No vs. medium	-9.24 %	-4.03	<.0001	(-13.8, -4.67)
No vs. high	-4.42 %	-1.89	.06	(-9.08, 0.25)
Medium vs. high	4.82 %	2.04	.045	(0.11, 9.53)

While the wait time and utilization results were expected due to the decrease in interaction times and longer neglect times with increasing heterogeneity, the performance score results in Fig. 4 showed a different trend, in that the medium heterogeneity configuration resulted in significantly higher score than both the no and high heterogeneity team structures. These results suggest that even though increasing the variety of the vehicles under control with different capabilities can reduce operator workload and system delays if NTs are increased, the use of higher levels of autonomy can also lead

to degraded performance. While this is an important finding that no doubt has significant implications, we leave this as an area of future research since the focus of this work is to develop a DES model that can both replicate these results and predict the likely outcomes of other team configurations.

B. DES Replications without Workload Effects

Using the participant data from the experiment, DES models were constructed for the three vehicle heterogeneity conditions. In RESCHU there were four different vehicle event types which required user interaction: 1) a vehicle arriving to an AOI and requiring the operator to undertake a search task (a vehicle-generated endogenous event), 2) an opportunity for re-planning the vehicle's path to a closer AOI (an operator-induced endogenous event), 3) an idle vehicle that requires assignment to an AOI (a vehicle-generated endogenous event), and 4) the intersection of a vehicle's path with a threat area (an exogenous environmental event). Table 2 presents the fitted distribution types and their parameters for these four different event arrivals and services. All distributions were generated from experimental data using distribution fitting software, assessed via Kolmogorov-Smirnov goodness-of-fit tests. Using these distributions, 10,000 trials were conducted for each DES model.

The probabilistic distribution parameters presented in Table 2 constitute the complete list of parameters used in the DES model. When replicating the high heterogeneity team structure, the model used the parameter estimates obtained from the high heterogeneity experimental data (column 4 in Table 2). Similarly, the medium heterogeneity and no heterogeneity conditions were replicated using the parameters obtained from the medium (column 5) and no (column 6) heterogeneity data, respectively.

As shown in Fig. 4, the model estimates for the three dependent variables do not fall in the 95% confidence intervals obtained from the observed data for both the search task wait times and operator utilization. Given that there were 10,000 trials run for the DES model, the standard errors for the estimated model means were practically 0. Thus, when the estimated means fall outside the 95% confidence intervals of the observed means, there is a statistically significant difference between the two. Therefore, this model fails to accurately replicate the observed data, which means that the model would also not be able to accurately predict for other UV team combinations. The following sections attempt to improve model replication by incorporating the workload effects on operator attention inefficiencies.

VI. UTILIZATION AND ATTENTION INEFFICIENCIES

As discussed previously, operator utilization, our measure of workload, is hypothesized to affect performance, such that it is degraded at both high and low ends of the utilization curve. In particular, utilization can guide how well operators notice events, inducing unnecessary wait times for vehicle servicing, in particular through attention switching delays (i.e., WTAI).

TABLE II
EVENT ARRIVAL AND SERVICE DISTRIBUTIONS FOR THE THREE HETEROGENEITY VEHICLE STRUCTURES

Vehicle	Event type	Event generator	Distribution (parameters)		
			Exp: Exponential, Γ : Gamma, Log-N: Log-normal, N: Normal, NA: Not Applicable		
			High Heterogeneity	Medium Heterogeneity	No Heterogeneity
MALE UAV	Type1	Search task arrival	Γ (α : 4.61, β : 21.97)	Γ (α : 4.04, β : 26.80)	Γ (α : 4.08, β : 26.27)
		Modified search task arrival due to re-plan	Log-N (μ : 3.86, σ : 0.54)	Γ (α : 3.21, β : 19.67)	Γ (α : 2.80, β : 18.65)
	Type2	Search task service	Log-N (μ : 3.14, σ : 0.59)	Log-N (μ : 2.96, σ : 0.64)	Log-N (μ : 2.94, σ : 0.63)
		Re-plan ratio	Bernoulli (p : .58)	Bernoulli (p : .51)	Bernoulli (p : 0.48)
		Re-plan service	N (μ : 3.19, σ : 7.32)	Exp (λ : 2.52)	Exp (λ : 3.2)
	Type3	Idle ratio	Bernoulli (p : .1)	Bernoulli (p : 0)	Bernoulli (p : 0)
		Idle duration	Exp (λ : 34.88)	NA	NA
		Idle service	N (μ : 3.19, σ : 7.32)	NA	NA
	Type4	Threat area arrival	Exp (λ : 105.49)	Exp (λ : 95.15)	Exp (λ : 72.9)
		Threat area service	Log-N (μ : 0.75, σ : 0.56)	Γ (α : 1.64, β : 1.49)	Γ (α : 0.58, β : 0.56)
UUV	Type1	Search task arrival	Γ (α : 2.89, β : 73.51)	N (μ : 174.07, σ : 102.33)	
		Modified search task arrival due to re-plan	Log-N (μ : 4.73, σ : 0.58)	Log-N (μ : 4.61, σ : 0.37)	
	Type2	Search task service	Log-N (μ : 2.95, σ : 0.69)	Log-N (μ : 2.78, σ : 0.66)	
		Re-plan ratio	Bernoulli (p : 1)	Bernoulli (p : 1)	
		Re-plan service	N (μ : 3.19, σ : 7.32)	Exp (λ : 2.52)	
	Type3	Idle ratio	Bernoulli (p : .69)	Bernoulli (p : .42)	
		Idle duration	Exp (λ : 35.91)	Exp (λ : 59.1)	
		Idle service	N (μ : 3.19, σ : 7.32)	Exp (λ : 2.52)	
	Type4	Threat area arrival	Exp (λ : 182.8)	Exp (λ : 168.63)	
		Threat area service	Log-N (μ : 0.75, σ : 0.56)	Γ (α : 1.64, β : 1.49)	
HALE UAV	Type1	Search task arrival	N (μ : 154.38, σ : 56.05)		
		Modified search task arrival due to re-plan	N (μ : 94.42, σ : 39.57)		
	Type2	Search task service	N (μ : 0.1, σ : 0.1)		
		Re-plan ratio	Bernoulli (p : .38)		
		Re-plan service	N (μ : 3.19, σ : 7.32)		
	Type4	Threat area arrival	Exp (λ : 151)		
		Threat area service	Log-N (μ : 0.75, σ : 0.56)		

In the experiment, WTAI was measured as the time from an emergent threat area intersection with a vehicle's path to the time when the participant responded to this intersection. The response to emergent threat areas was chosen as the measure of WTAI since avoiding threat areas was the highest priority task for the participants, and it required decisive and identifiable actions.

The experimental data revealed that the overall average post-hoc utilization values for the different vehicle heterogeneity levels ranged between 40 and 80%. However, this static post-hoc calculation does not reflect the dynamic nature of utilization, so four utilization values were calculated for 2.5 minutes time windows for the 10 minute experiment. The average WTAI for different values of UT across the four time intervals are presented in Fig. 5a. Due to missing data, the number and spread of utilization bins for each condition differed. In the case of the no heterogeneity condition, only four bins had enough samples, all at higher utilization values due to the high operational tempo. Fig. 5b demonstrates the associated performance scores for these same utilization bins.

In order to determine if significant differences existed in WTAI across 10% utilization bins for the three different

heterogeneity conditions, a repeated measures Analysis of Variance (ANOVA) was conducted. A logarithmic transformation was performed on WTAI to meet statistical modeling assumptions, that is, normality and homogeneity of variances. Results revealed that there were significant differences between different utilization intervals for the medium ($F(5,119)=2.43$, $p=.04$) and high heterogeneity conditions ($F(5,121)=8.11$, $p<.0001$). Differences in WTAI for the low heterogeneity condition was only marginally significant ($F(3,110)=2.54$, $p=.06$).

In the no heterogeneity case, pair-wise comparisons showed that 60-70% utilization resulted in shorter WTAI than both the 80-90% ($p=.03$) and 70-80% utilization bins ($p=.02$). In the medium heterogeneity case, the 80-90% utilization bin resulted in longer WTAI than for the 60-70% bin ($p=.03$), 50-60% bin ($p=.008$), and 40-50% bin ($p=.04$). In addition, the 30-40% utilization bin also resulted in longer WTAI than the 60-70% bin ($p=.047$) and the 50-60% bin ($p=.02$). Finally, in the high heterogeneity case, 80-90% utilization bin resulted in longest WTAI when compared to all other utilization values (70-80%: $p=.04$; 60-70%: $p=.001$; 50-60%: $p<.0001$; 40-50%: $p<.0001$; 30-40%: $p<.0001$). However, the 30-40% utilization

resulted in significantly shorter WTAI when compared to 60-70% ($p=.002$), and 70-80% utilization ($p=.005$).

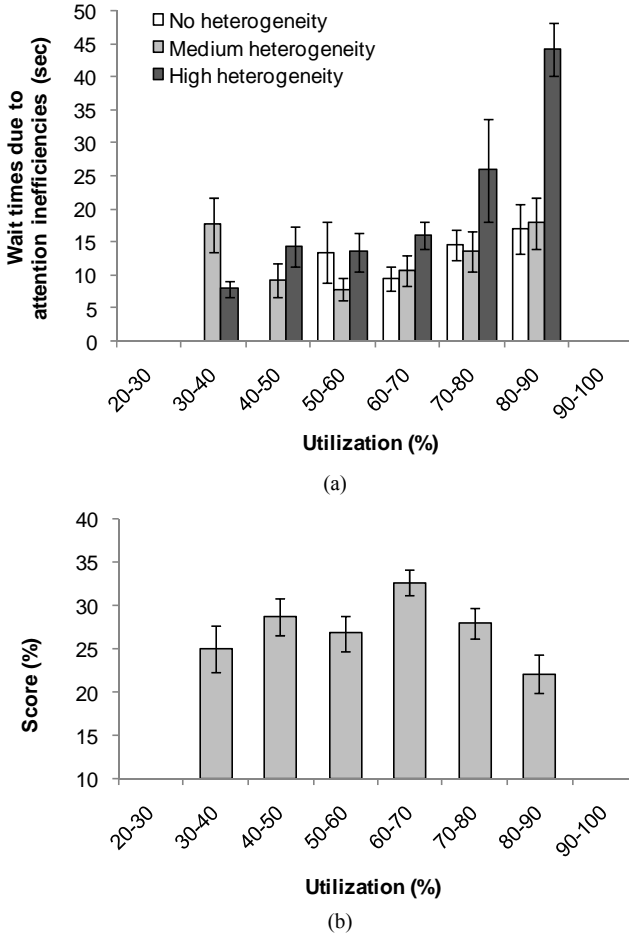


Fig 5. (a) Experimental results for UT/WTAI relation (with standard error bars) (b) Experimental results for score vs. UT (with standard error bars).

These results demonstrate that WTAI is longer at higher utilization levels than at medium utilization levels, consistent with our initial hypothesis. However, the medium and high heterogeneity conditions contradict in terms of how utilization is related to WTAI for low utilization levels. While the medium condition is in agreement with the initial hypothesis that lower utilization values have higher WTAI than medium utilization values, the high heterogeneity case resulted in the reverse trend. As most of the data points for all three conditions fell towards higher utilization values and the 10 minute experiment did not require vigilance to be maintained for a long period of time, the functional form of WTAI at lower utilization values is left for future work. Moreover, at high utilization values, WTAI appears to be much higher for the high heterogeneity vehicle structure. This suggests that high operator utilization resulting from controlling multiple vehicles with different capabilities can be especially detrimental to WTAI, and thus overall mission performance.

In order to assess if the UT/WTAI relation was consistent across the four different time windows, another repeated measures ANOVA was conducted using the experimental data

from all three vehicle heterogeneity levels. Vehicle heterogeneity could not be included as a factor in the model due to the large number of missing design cell combinations, and the small number of observations in some of the design cells. A logarithmic transformation was performed on WTAI to stabilize variance. The time window ($p=.26$) and the time window-UT interaction were not significant ($p=.71$), suggesting that UT/WTAI relation can be considered as fairly consistent across the different time windows.

The distribution of the performance scores (Fig. 5b) suggests that over or under utilization caused overall mission performance to degrade. A mixed linear regression model demonstrated that utilization was significantly associated with score ($F(5,126)=4.2$, $p=.001$), given a backward selection model, controlling for vehicle heterogeneity level ($F(2,71)=8.37$, $p=.0005$), time window ($F(3,208)=24.30$, $p<.0001$), and vehicle heterogeneity level-time window interaction ($F(6,208)=2.08$, $p=.06$).

Pair-wise comparisons for Fig. 5b revealed that 60-70% utilization corresponded to significantly higher scores than most other utilization values (30-40%: $p=.02$, 50-60%: $p=.02$, 70-80%: $p=.02$, 80-90%: $p<.0001$). In addition, 80-90% utilization resulted in lower scores than both 70-80% ($p=.03$) and 40-50%, ($p=.04$) utilization. Previous studies have also shown that when the operators work beyond 70% utilization, performance degrades significantly [19, 22], so these results also demonstrate that there is a threshold for performance in terms of operator utilization.

VII. DES REPLICATIONS WITH MODELING WORKLOAD EFFECTS

The experimental data previously described was used to incorporate the effects of workload in the DES model. This revision of the DES model included modification of the arrival of events (both exogenous and endogenous) so that events arrive to the system once they are noticed by the operator. Thus, an operator's inattention efficiency due to workload was modeled stochastically by introducing vehicle wait times as a function of utilization. Whenever there was an event arrival, the utilization for the previous 2.5 minute time window was calculated, which in turn was used to identify the appropriate time penalty from the UT/WTAI relation as presented in Fig. 5a.

As shown in Fig. 4, the revised DES model estimates for the three dependent variables (i.e., search task wait times, utilization, and score) fall in the 95% confidence intervals obtained from the observed data. Therefore, the revised DES model, which accounts for WTAI, more accurately estimates the observed data for all vehicle heterogeneity levels, especially when compared to the results without including this relationship. This suggests that WTAI can be inserted into a DES model as a function of operator utilization, which can be fed back through the model, providing a better estimate of the operator's influence on the system.

VIII. DES PREDICTIONS WITH MODELING WORKLOAD EFFECTS

This section presents the effects of WTAI on the model's predictive power, that is, the model's ability to predict the effects of changes in the vehicle heterogeneity structure. In order to assess predictive power, the DES model was constructed using the experimental data for the medium heterogeneity condition in order to predict for the no and high heterogeneity levels. Therefore, model distributions (e.g., service times and UT/WTAI relation) were populated based on this experimental data subset only. The medium heterogeneity condition was chosen to build the model because we wanted to assess model's predictive capabilities for both increasing and decreasing heterogeneity.

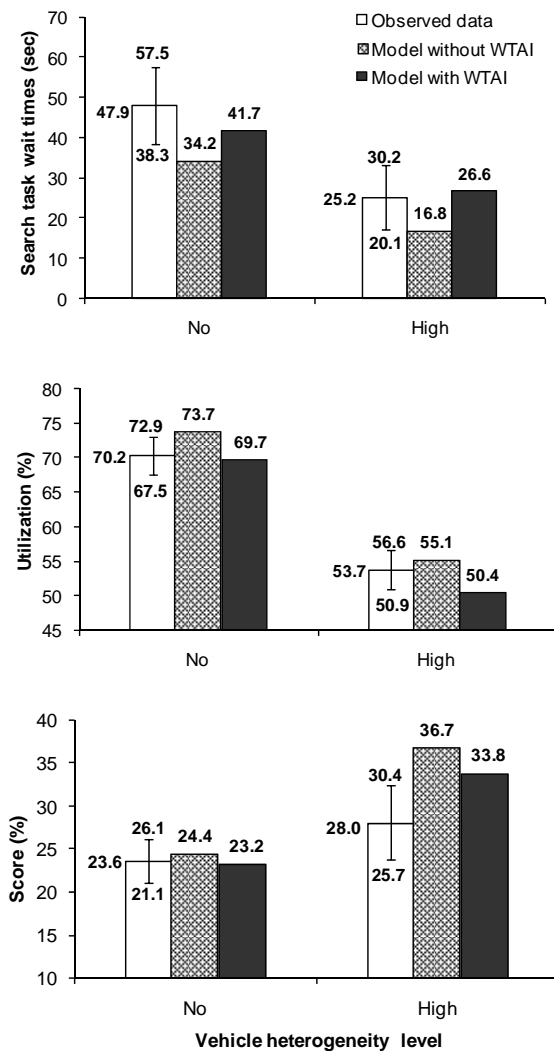


Fig 6. PREDICTION: Observed data with 95% confidence intervals and model predictions for the three dependent variables of interest (a) search task wait times, (b) utilization, and (c) score.

In going from the medium to the no heterogeneity team, the change required replacement of two UUVs by two MALE UAVs. Thus in the DES, the appropriate arrival and service

processes were substituted. The parameter estimates for MALE UAVs from column 5 in Table 2 were used to predict for the no heterogeneity vehicle team structure.

In going from the medium heterogeneity team to the high heterogeneity team, one of the MALE UAVs was replaced by a HALE UAV. The parameter estimates for MALE UAVs and UUVs from column 5 in Table 2 were used when predicting for the high heterogeneity vehicle team structure, in particular when modeling MALE UAVs and UUVs. However, simple arrival distribution substitution was not possible for HALE UAVs because there were no HALE UAVs in the medium heterogeneity team and therefore the arrival processes of vehicle-generated events could not be derived for this type of vehicle. A Monte Carlo simulation was used to derive the missing data. More specifically, Monte Carlo simulations were used to derive travel times between randomly located AOIs, which translated to a new vehicle-generated event arrival. The samples from the simulations were then used to build the arrival distributions for this condition. The service times for HALE UAVs were assumed to be negligible, since servicing HALE UAVs only required clicking on a button which took much shorter than servicing MALE UAVs or UUVs.

As discussed previously, the degree of heterogeneity in team structure resulted in significant differences in operator utilization, search task wait times, and score. The DES model incorporating WTAI more accurately predicted these observed changes in search task wait times and the operator utilization than did the DES model that did not account for workload (Fig. 6). In the case of the performance score and utilization, the predictions were accurate for the homogeneous condition, but for the high vehicle heterogeneity, the revised model's estimates were not as accurate. This inaccuracy is likely due to the missing data problem. However, the DES model still captured the trend of increasing performance score as the team structure changed from no heterogeneity to high.

IX. CONCLUSIONS

This paper demonstrates the incorporation of the effects of workload in a discrete event simulation model of human supervisory control, in particular, multiple unmanned vehicle supervisory control. As demonstrated in a human-in-the-loop experiment, system delays caused by operator attention inefficiencies (measured through attention switching delays) are significantly related to operator utilization, and these system delays can negatively impact the overall mission. High levels of workload (measured through utilization), and in some cases low workload, led to increased attention switching delays. Consequently system wait times increased, which ultimately led to poor mission performance. The DES model with the inclusion of these additional wait times due to operator attention inefficiencies provided enhanced accuracy for replicating experimental observations and predicting results for different UV team structures, as compared to the DES model without accounting for operator workload.

It is important to note that this DES model does not capture all possible human cognitive inefficiencies, but rather an

aggregate effect of inefficient information processing, which likely has many sources that exist at a level difficult to capture in a DES model. There are likely many more sources of cognitive inefficiencies, such as operator trust [39], that are manifested through system delays. However, in the spirit of Occam's Razor, we have focused on a single quantifiable relationship that provides bounded estimates of operator behavior.

One issue that requires further investigation is the temporal and dynamic nature of utilization. All utilization metrics reported here were post-hoc aggregate measures, which are not accurate as an instantaneous measure of workload. Thus more research is needed to determine how utilizations can be measured in a real-time fashion, and moreover, what thresholds truly indicate poor performance, i.e., does a person need to work at or above 70% utilization for some period of time before the negative effects are seen in human and/or system performance?

Another related area that requires further investigation is the rate of onset of high workload or utilization periods. The true measure of the impact of workload on performance may not be sustained utilization, but rather the onset rates of increasing utilizations. The effects of low utilization or cognitive underload, and the nature of sustained and variable rates of utilization changes also deserve further scrutiny.

Such investigations will be crucial in both aiding supervisory control modeling efforts, but they are also potentially valuable in the field of dynamic, adaptive automation design. If successful performance models of over or under cognitive load based on utilization can be developed, then more reliable forms of adaptive automation can be developed such that automation can intervene or assist human operators when a transition into a negative workload-performance region occurs.

The methodology used in this paper presents a way to dynamically adjust performance scores based on operator workload in order to make generally effective system predictions via DES models. Since the discrete event simulation model was substantially improved with the consideration of the effects of workload, this research has implications towards developing more realistic models of human supervisory control and human-system performance.

ACKNOWLEDGMENTS

The authors would like to thank Yale Song for his help in coding the experimental test-bed and Paul de Jong for his help in data processing.

REFERENCES

- [1] T. B. Sheridan, *Telerobotics, Automation, and Human Supervisory Control*. Cambridge, MA: The MIT Press, 1992.
- [2] DOD, "Unmanned Systems Roadmap," Office of the Secretary of Defense, Washington, D.C. 2007-2032, 2007.
- [3] D. S. Alberts, J. J. Garstka, and F. P. Stein, *Network Centric Warfare: Developing and Leveraging Information Superiority*, 2nd ed.: CCRP Publication Series, 1999.
- [4] A. Feickert, "The Army's Future Combat System (FCS): Background Issues for Congress," Congressional Research Service RL32888, 2008.
- [5] M. L. Cummings, C. E. Nehme, J. W. Crandall, and P. J. Mitchell, "Predicting operator capacity for supervisory control of multiple UAVs," in *Innovations in Intelligent Machines*, J. S. Chahl, L. C. Jain, A. Mizutani, and M. Sato-Ilic, Eds., 2007, pp. 11-36.
- [6] M. L. Cummings, S. Bruni, S. Mercier, and P. J. Mitchell, "Automation architecture for single operator, multiple UAV command and control," *The International Command and Control Journal*, vol. 1, pp. 1-24, 2007.
- [7] A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, 3rd ed. Boston: McGraw-Hill International Series, 2000.
- [8] M. Woolridge, *Introduction to Multiagent Systems*, 2nd ed. New York, NY: John Wiley & Sons, Inc., 2009.
- [9] Y. Shoham and K. Leyton-Brown, *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge: Cambridge University Press, 2009.
- [10] J. W. Crandall, M. A. Goodrich, R. O. Olsen, and C. W. Nielsen, "Validating human-robot interaction schemes in multitasking environments," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 35, pp. 438-449, 2005.
- [11] S. Mau and J. Dolan, "Scheduling to minimize downtime in human-multirobot supervisory control," in *Proceedings of the International Workshop on Planning and Scheduling for Space*, 2006.
- [12] R. O. Olsen and M. A. Goodrich, "Metrics for evaluating human-robot interactions," in *Proceedings of NIST Performance Metrics for Intelligent Systems Workshop*, 2003.
- [13] M. L. Cummings and P. J. Mitchell, "Predicting controller capacity in supervisory control of multiple UAVs," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 38, pp. 451-460, 2008.
- [14] J. W. Senders, "The human operator as a monitor and controller of multidegree of freedom systems," *IEEE Transactions on Human Factors in Electronics*, vol. 5, pp. 2-5, 1964.
- [15] T. B. Sheridan, "On how often the supervisor should sample," *IEEE Transactions on Systems, Science, and Cybernetics*, vol. 6, pp. 140-145, 1969.
- [16] J. R. Carbonell, J. L. Ward, and J. W. Senders, "A queuing model of visual sampling experimental validation," *IEEE Transactions on Man Machine Systems*, vol. 9, pp. 82-87, 1968.
- [17] Y. Liu, "Queuing network modeling of human performance of concurrent spatial and verbal tasks," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 27, pp. 195-207, 1997.
- [18] Y. Liu, R. Feyen, and O. Tshimhoni, "Queuing network-model human processor (QN-MHP): a computational architecture for multitask performance in human-machine systems," *ACM Transactions on Computer-Human Interaction*, vol. 13, pp. 37-70, 2006.
- [19] D. K. Schmidt, "A queuing analysis of the air traffic controller's workload," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, pp. 492-498, 1978.
- [20] C. D. Wickens and J. G. Hollands, *Engineering Psychology and Human Performance*, 3rd ed Upper Saddle River, NJ: Prentice-Hall, 2000.
- [21] W. W. Wierwille and F. T. Eggemeier, "Recommendations for mental workload measurement in a test and evaluation environment," *Human Factors*, vol. 35, pp. 263-281, 1993.
- [22] M. L. Cummings and S. Guerlain, "Developing operator capacity estimates for supervisory control of autonomous vehicles," *Human Factors*, vol. 49, pp. 1-15, 2007.
- [23] W. B. Rouse, *Systems Engineering Models of Human-Machine Interaction*. New York: North Holland, 1983.
- [24] D. Kahneman, *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall, 1973.
- [25] D. O. Hebb, "Drives and the C. N. S. (conceptual nervous system)," *Psychological Review*, vol. 62, pp. 243-254, 1955.
- [26] R. M. Yerkes and J. D. Dodson, "The relation of strength of stimulus to rapidity of habit-formation," *Journal of Comparative Neurology and Psychology*, vol. 18, pp. 459-482, 1908.
- [27] R. W. Proctor and T. Van Zandt, *Human Factors in Simple and Complex Systems*, 2nd ed. Boca Raton, FL: CRC Press, 2008.
- [28] R. Parasuraman and D. R. Davies, "Decision theory analysis of reponse latencies in vigilance," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 2, pp. 569-582, 1976.

- [29] C. Wu and Y. Liu, "Queuing network modeling of driver workload and performance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, pp. 528-537, 2007.
- [30] C. Wu, O. Tsimhoni, and Y. Liu, "Development of an adaptive workload management system using the queueing network-model human processor (QN-MHP)," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, pp. 463-475, 2008.
- [31] C. Nehme, B. Mekdeci, J. W. Crandall, and M. L. Cummings, "The impact of heterogeneity on operator performance in futuristic unmanned vehicle systems," *The International C2 Journal*, vol. 2, pp. 1-30, 2008.
- [32] J. W. Crandall and M. L. Cummings, "Identifying predictive metrics for supervisory control of multiple robots," *IEEE Transactions on Robotics - Special Issue on Human-Robot Interaction*, vol. 23, pp. 942-951, 2007.
- [33] M. Pinedo, *Scheduling: Theory, Algorithms, and Systems*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 2002.
- [34] E. D. Lazowska, J. Zahorjan, G. S. Graham, and K. C. Sevcik, *Quantitative System Performance: Computer System Analysis Using Queueing Network Models*. Upper Saddle River, NJ: Prentice Hall, Inc, 1984.
- [35] D. E. Broadbent, *Perception and Communication*. Oxford: Pergamon, 1958.
- [36] A. T. Welford, "The psychological refractory period and the timing of high-speed performance - a review and a theory," *British Journal of Psychology*, vol. 43, pp. 2-19, 1952.
- [37] C. Wu, Y. Liu, and C. M. Quinn-Walsh, "Queuing network modeling of a real-time psychophysiological index of mental workload-P300 in event-related potential (ERP)," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 38, pp. 1068-1084, 2008.
- [38] C. Nehme, "Modeling human supervisory control in heterogeneous unmanned vehicle systems." vol. Ph.D. dissertation Cambridge, MA: Massachusetts Institute of Technology, 2009.
- [39] J. D. Lee and K. A. See, "Trust in automation: designing for appropriate reliance," *Human Factors*, vol. 46, pp. 50-80, 2004.

Birsen Donmez received her B.S. in mechanical engineering from Bogazici University in 2001, her M.S. (2004) and Ph.D. (2007) in industrial engineering, and her M.S. in statistics (2007) from the University of Iowa.

She is currently an assistant professor in the Mechanical & Industrial Engineering Department at the University of Toronto. Before joining University of Toronto, Birsen was a postdoctoral associate at the Massachusetts Institute of Technology. Her research interests are centered on understanding and improving human behavior and performance in multi-task and complex situations, using a wide range of analytical techniques.

Carl Nehme received a Bachelor in computer engineering from the University of Toronto in 2002, his M.S. (2004) and Ph.D. (2008) in aeronautics and astronautics from the Massachusetts Institute of Technology.

Carl is currently a consultant at McKinsey & Company.

Mary L. Cummings (M'03) received her B.S. in mathematics from the United States Naval Academy in 1988, her M.S. in space systems engineering from the Naval Postgraduate School in 1994, and her Ph.D. in systems engineering from the University of Virginia in 2003.

A naval officer and military pilot from 1988-1999, she was one of the Navy's first female fighter pilots. She is currently an associate professor in the Aeronautics & Astronautics Department at the Massachusetts Institute of Technology. Her previous teaching experience includes instructing for the U.S. Navy at Pennsylvania State University and as an assistant professor for the Virginia Tech Engineering Fundamentals Division. Her research interests include human supervisory control, human-uninhabited vehicle interaction, bounded collaborative human-computer decision making, decision support, information complexity in displays, and the ethical and social impact of technology.